CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Self-Affirmation and Identity-Driven Political Behavior

Benjamin A. Lyons[1],* (iD), Christina E. Farhart[2], Michael P. Hall[3], John Kotcher[4], Matthew Levendusky[5], Joanne M. Miller[6], Brendan Nyhan[7], Kaitlin T. Raimi[8], Jason Reifler[9], Kyle L. Saunders[10], Rasmus Skytte[11] and Xiaoquan Zhao[12]

[1]Department of Communication, University of Utah, Salt Lake City, UT, USA, Twitter: @ba_lyons, [2]Department of Political Science, Carleton College, Northfield, MN, USA, Twitter: @CFarhart, [3]Department of Psychology, University of Michigan, Ann Arbor, MI, USA, [4]Center for Climate Change Communication, George Mason University, Fairfax, VA, USA, Twitter: @jekotcher, [5]Department of Political Science, University of Pennsylvania, Philadelphia, PA, USA, Twitter: @m_levendusky, [6]Department of Political Science & International Relations, University of Delaware, Newark, DE, USA, Twitter: @PoliPsyProf, [7]Department of Government, Dartmouth College, Hanover, NH, USA, Twitter: @BrendanNyhan, [8]Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor, MI, USA, Twitter: @KaitlinRaimi, [9]Department of Politics, University of Exeter, Exeter, UK, Twitter: @JasonReifler, [10]Department of Political Science, Colorado State University, Fort Collins, CO, USA, Twitter: @KyleLSaunders, [11]Department of Political Science, Aarhus University, Aarhus, Denmark, Twitter: @RasmusSkytte and [12]Department of Communication, George Mason University, Fairfax, VA, USA
*Corresponding author. Email: Benjamin.a.lyons@gmail.com

Abstract

Psychological attachment to political parties can bias people's attitudes, beliefs, and group evaluations. Studies from psychology suggest that self-affirmation theory may ameliorate this problem in the domain of politics on a variety of outcome measures. We report a series of studies conducted by separate research teams that examine whether a self-affirmation intervention affects a variety of outcomes, including political or policy attitudes, factual beliefs, conspiracy beliefs, affective polarization, and evaluations of news sources. The different research teams use a variety of self-affirmation interventions, research designs, and outcomes. Despite these differences, the research teams consistently find that self-affirmation treatments have little effect. These findings suggest considerable caution is warranted for researchers who wish to apply the self-affirmation framework to studies that investigate political attitudes and beliefs. By presenting the "null results" of separate

research teams, we hope to spark a discussion about whether and how the self-affirmation paradigm should be applied to political topics.

Keywords: Self-affirmation; polarization; partisanship; political behavior

Psychological attachment to political parties and related identities and values can bias information processing, belief and attitude formation, and group evaluations. Researchers have sought to apply self-affirmation theory (Steele, 1988) to this problem. Self-affirmation theory proposes that the overall goal of the self is to protect one's view of their own self-integrity. In response to threats to this view, people act to restore self-worth through defensive reactions. Alternatively, though, self-worth can be restored through an affirmation of other sources of self-integrity, such as one's commitment to personally important values. This alternative route to protected self-integrity in the face of threatening information can reduce the need to rely on defensive biases

Several studies report promising results in explicitly partisan contexts (Badea et al. 2017; Badea et al. 2018; Binning et al. 2010; Binning et al. 2015; Carnahan et al. 2018; Cohen et al. 2007; Van Prooijen and Sparks, 2014), which they interpret as demonstrating that affirming people's self-worth in nonpolitical domains might serve to reduce defensive or biased behavior resulting from a perceived risk to one's identity or self-concept. This general finding – that self-affirmation may ameliorate partisan biases – has since been canonized in reviews (Cohen and Sherman, 2014; Sherman and Cohen, 2006). While the broader self-affirmation literature in psychology is voluminous, recent studies have called into question the uniformity of its effects (e.g. Reavis et al. 2017). Large-scale replications (Hanselman et al. 2017; Protzko and Aronson, 2016) suggest that self-affirmation interventions may be "fragile" in some domains of social psychology (but see Borman et al. 2018).

In this article, we report a series of studies conducted independently by separate research teams that examine self-affirmation in political contexts. We employ conceptual extensions of prior work applying self-affirmation to politics, rather than close replications. Indeed, our goal was to add more evidence about previous self-affirmation applications' effectiveness in new, but theoretically related outcomes in politics. In other words, we apply self-affirmation procedures to new political outcomes of interest, and note how and if we depart from any prior work that found significant effects. Our results find that self-affirmation treatments consistently have a little effect across a range of samples and outcome measures (attitudes, factual beliefs, conspiracy beliefs, affective polarization, and evaluations of news sources).[1] Our hope is that these findings help unite evidence about the study of self-affirmation in political contexts. The studies we report suggest that self-affirmation may have more limited potential in political contexts than previously thought. We hope that future scholars considering using self-affirmation in a political context will consider the evidence about the prior literature we collect here, as well as the variety of interventions and designs we report, and that these results will be useful to future work considering self-affirmation's potential to reduce partisan biases.

---

[1]In three instances, data presented here have been reported as part of broader research projects (Levendusky, 2018; Lyons, 2018; Nyhan and Reifler, 2019).

Self-affirmation theory

The theory of self-affirmation is based on the premise that people resist threats to their sense-of-self (e.g. processing counter-attitudinal political information in a biased manner) *and* that self-integrity can cross domains. In other words, defensive processing in one aspect of a person's self-concept may be tempered by bolstering another (Steele, 1988). According to this account, threats in one domain lose their potency when self-worth is affirmed in other core aspects of one's self-worth, eliminating the need for ego protection. This elegant theoretical account explains a general problem (biased processing), a mechanism (ego-defending reactions in response to threat), and a potential solution (debiasing by affirming one's sense-of-self in an unrelated domain). Self-affirmation is canonically induced by way of writing prompts that ask respondents to reflect on important values or characteristics they hold, and describe experiences in which these were exhibited or played an important role in their life.

The core tenet of the self-affirmation approach is that threats to highly central or salient social identities could result in significant "costs," which promotes defensive responses absent alternative sources of self-integrity (Cohen et al. 2007; Sherman and Cohen, 2006, p. 218). Previous literature suggests that threats to social identity (such as one's political affiliation) can be buffered through affirmation, which could lead to less anchoring of group evaluations in one's self-concept and thereby allowing individuals to evaluate groups independent of their self-evaluation (Cohen et al. 2000). The importance of the domain, or particular social group, is supposed to condition the degree of self-threat and therefore the effectiveness of affirmation (Boninger et al. 1995). Consequently, any given self-affirmation intervention is not predicted to affect everyone equally. Rather, the effect should be conditional or contingent for those who are being confronted with a threat to their identity. Consequently, each of the studies we present here examines the effect of self-affirmation on the appropriate target population.

An important qualifying condition for the relevance of the self-affirmation approach is to establish that political identities are available to people. Klar (2013) shows that partisan identity can be made salient through mere mention. Asking questions about respondents' feelings toward their own party and the opposing party, as would commonly occur in a survey about political topics, makes partisan identity salient. Threat is an even more powerful means of raising salience (Klar, 2013). For strong partisans or those with strong prior attitudes, many of the experimental setups themselves constitute a threat. For example, those with strong beliefs against the scientific consensus on climate change view contrary information as a threat (Ma et al. 2019). Partisans sometimes see the opposing party as a threat to their way of life (Iyengar et al. 2019). Mason (2018) compellingly shows the centrality of partisanship to identity and the extent to which this generates powerful negative affect towards political outgroups, creating clear opportunities to experience the type of threat that self-affirmation is theorized to protect against. In sum, for strong partisans, questions about political beliefs and feelings toward political groups themselves generate reactance and negative emotion. Under these conditions, former work predicts we should see

Table 1a
Summary of Unpublished Studies

Authors	Self-affirmation approach	Subject matter	Dependent variable(s)	Sample	Year
Farhart et al.	SA without information treatment	Ideological conspiracy endorsement	Ideologically consistent conspiracy theory indices	MTurk ($N = 3799$)	2016
Kotcher	SA without information treatment	Climate change	Climate change attitudes and beliefs	Toluna ($N = 696$)	2016
Raimi and Hall	SA without information treatment	Immigration	Belief extremity; belief superiority	MTurk ($N = 400$)	2016
Skytte	SA without information treatment	Affective polarization	Party affect; candidate affect; outparty tolerance	MTurk ($N = 204$)	2016
Zhao	Full cross of SA and information treatment	Climate change	News evaluations	Other ($N = 455$)	2013

Table 1b
Summary of Studies Published During Project

Authors	Self-affirmation approach	Subject matter	Dependent variable(s)	Sample	Year
Levendusky	SA without information treatment	Affective polarization	Party affect; outparty threat; social distance	GfK [TESS] ($N = 1{,}345$)	2016
Lyons Study 1	Information following randomized SA	Mining spills	Attribution of blame and motivation	MTurk ($N = 130$)	2015
Lyons Study 2	Information following randomized SA	Biofuels	Risk/benefit perception	MTurk ($N = 274$)	2016
Nyhan and Reifler Study 1	Full cross of SA and information treatment	Foreign policy (USA/Iraq)	Trend/change in insurgent attacks in Iraq	YouGov [CCES] ($N = 525$)	2008
Nyhan and Reifler Study 2	Full cross of SA and information treatment	Economy	Trend/change in number of jobs 2010–2011	MTurk ($N = 247$)	2011
Nyhan and Reifler Study 3	Full cross of SA and information treatment	Climate change	Trend/change in global average surface temperature	Qualtrics ($N = 244$)	2011

self-affirmation reduce negative attitudes or group conformity (see, e.g. Van Prooijen and Sparks, 2014).

A summary of each study's approach can be found in Tables 1a and 1b. Further details can be found in the Supplementary Materials.

Selection of studies

The corresponding author solicited "file drawer" studies examining self-affirmation and political behavior, regardless of findings, in March 2018. No studies nominated in the process were excluded. All the authors then shared data allowing for a parallel analysis, as described below. Over the course of the project, three research teams published articles based on the data provided to the corresponding author. (We report all findings from the different research teams using the same analytic strategy. The results we report here for studies that were accepted during this process are substantively consistent with the published version, but the model specifications and point estimates are different.) To aid the reader, we have created separate tables to distinguish between published and unpublished studies.

We report results from a number of studies that examine various outcomes in which political identity may drive biases. Specifically, we look at self-affirmation's potential to mitigate conspiracy beliefs (e.g. Miller, Saunders, and Farhart, 2016; Oliver and Wood, 2014), affective polarization (e.g. Iyengar et al. 2019; Mason, 2018), belief superiority (e.g. Saucier and Webster, 2010), news evaluations (e.g. Pingree et al. 2014), and various forms of party-aligned factual beliefs (e.g. Kahan and Braman, 2006; Taber and Lodge, 2006).

Analytic approach

Social psychology experiments that examine the role that self-affirmation may have in reducing bias or extremity typically fall into three types. The first type of study looks at the effect of self-affirmation for relevant versus nonrelevant groups (Binning et al. 2010; Correll et al. 2004; Harris and Napper, 2005; Sherman et al. 2000; Van Koningsbruggen et al. 2009). For example, a message detailing coffee's health effects may be shown to drinkers and nondrinkers. A typical study would randomly assign subjects to a self-affirmation task (vs. a placebo task), and then show all participants the same stimulus. Analysis would then focus on the effects of self-affirmation for relevant versus nonrelevant groups. The Lyons studies follow this approach.

The second type of study examines the effect of self-affirmation on attitudes and extreme beliefs without an information treatment. These studies randomize assignment to a self-affirmation task (vs. a placebo task), and then measure outcomes of interest (e.g. Lehmiller et al. 2010). While not always framed in this way, from a political science perspective, one can see these studies as assessing whether self-affirmation makes people more amenable to considering counter-attitudinal information already encoded when constructing a survey response, a la Zaller (1992). The Farhart et. al., Kotcher, Levendusky, Raimi and Hall, and Skytte studies follow this approach.

Finally, some researchers examine how self-affirmation influences information processing in the face of a threat by concurrently manipulating both self-affirmation and information treatments. For example, Reavis et al. (2017) examine how affirmation moderates the effect of a threatening article correcting the MMR autism link (vs. a placebo article) on intent to vaccinate. Some of our studies follow this route (Nyhan and Reifler and Zhao studies).

As a result, we take two approaches to analysis. First, we look at each study as a two-cell experiment in which participants were randomly assigned to self-affirmation or a control treatment. We construct our analyses so that stimuli were either uniform across all participants or absent. While this approach easily incorporates studies that only manipulate self-affirmation, we need to take an extra step to incorporate studies that also manipulate information. When participants were shown one of the two information treatments (Zhao; Nyhan and Reifler Study 3), we split the study into two analyses – one for each information treatment. Thus, the analyses examine the effect of receiving being self-affirmed versus not being self-affirmed for those who received an information treatment. When the control group received no information (Nyhan and Reifler Studies 1 and 2), we only include cells where participants received the information treatment;[2] again comparing the effect of being self-affirmed versus not being self-affirmed for those who receive an information treatment. (The supplementary analyses include reanalyses of self-affirmation X information full crosses when excluded given the approach above.)

We use OLS regression with robust standard errors to estimate the effect of self-affirmation in each study. In each model, we estimate the effect of the affirmation treatment as an indicator variable, as well as a relevant moderator – such as party identity or strength of partisanship – and the corresponding interaction term identified by the authors. Interaction terms are constructed such that in most cases, they test the effect of self-affirmation in *attenuating* the effects of partisanship or ideology on a "negative" outcome.[3] To compare effects directly, we rescale all dependent variables to range from 0 to 1.

Treatment of multiple outcome measures

Some studies collected multiple relevant dependent variables. We deal with this in two ways. In the Kotcher, Skytte, and Zhao analyses, we create a composite item due to high internal consistency among measures, which also reduces measurement

---

[2]We drop from these analyses irrelevant conditions, such as alternate (non-self-affirmation) treatment arms (Levendusky, Lyons, and Raimi and Hall studies).

[3]Farhart et al. test self-affirmation's effect on the relationship between Democratic and Republican party affiliation and belief in party-congenial conspiracies. Levendusky and Skytte test its effect on the relationship between strength of party affiliation and affective polarization. Lyons tests its effect on the relationship between strength of party affiliation and group-aligned factual beliefs. Nyhan and Reifler test its effect on the relationships between a series of issue attitudes and partisanship on misperceptions. Raimi and Hall test its effect on the relationship between strength of party affiliation and belief superiority and extremity. Zhao tests its effect on the relationship between Republican party affiliation and perceived credibility of a "climate change is a hoax" message. However, Zhao also tests its effect on the relationship between Republican party affiliation and perceived credibility of a message suggesting climate change is legitimate (thus, a "positive" valence outcome), and similarly, Kotcher tests its effect on the negative relationship between conservative ideology and climate attitudes.
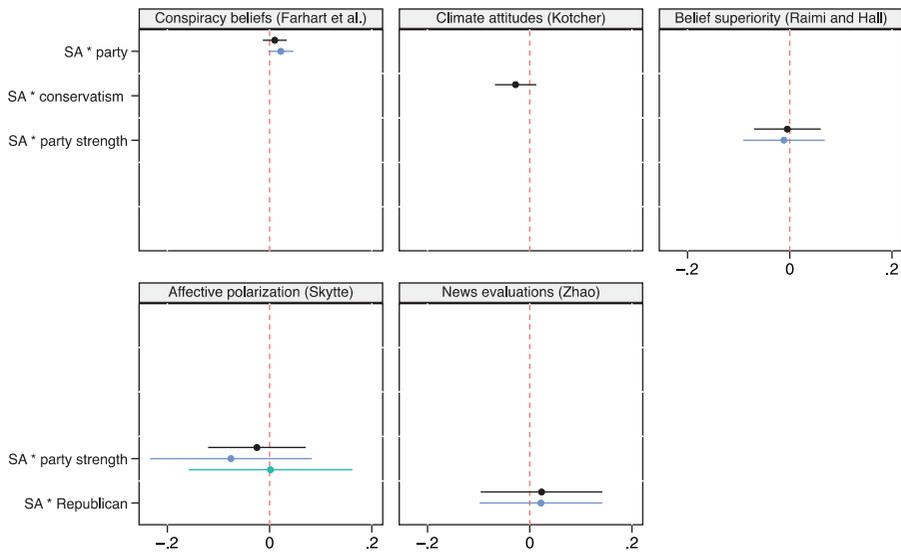
Figure 1
Interaction terms across unpublished studies.
NOTES: All outcome variables are rescaled to range from 0 to 1. Corresponding statistical information is shown in Tables 2a and 2b. For full regression models, see appendix.

error and improves power (Ansolabehere, Rodden, and Snyder, 2008). In cases where dependent variables are potentially orthogonal to one another, we analyze the effects of the self-affirmation treatment on each dependent variable separately.

Summary of results

We find no main effects across the studies. However, the theory of self-affirmation expects that effects should be limited to specific (threatened or counter-attitudinal) subgroups. Therefore, we relegate these analyses to the appendix. Our focus now is on the interaction between self-affirmation and an indicator variable identifying the relevant (threatened or counter-attitudinal) subgroups. In Figure 1, we plot regression coefficients for all interactions of the unpublished studies. We present interaction terms across both published and unpublished studies in Tables 2a and 2b.

The interaction term of the pooled unpublished studies is $b = 0.01$, SE $= 0.006$ (95% CI: $-0.004, 0.020$) (using a random-effects model and formula provided in Neyeloff et al. 2012), though we urge caution due to the differing outcome measures across these studies (Carpenter, 2020). We employ a random-effects model as we allow that the true effect size might differ from study to study, especially given the range of outcome measures. The interaction term when using all of the studies in the present manuscript (including those published) is $b = 0.00$, SE $= 0.004$ (95% CI: $-0.011, 0.006$).

Finally, when reanalyzing the full cross in experiments that included information treatments (Nyhan and Reifler Studies 1–3, and Zhao), we find effects similar to those above (see Tables A8–A9).

Table 2a
Coefficients for Self-Affirmation X Subgroup Indicator Interaction (unpublished)

Outcome measure	Sample size	B	SE	CI lower	CI upper
Liberal conspiracy beliefs (Farhart et al)	3796	0.02	0.01	0.00	0.05
Conservative conspiracy beliefs (Farhart et al)	3797	0.01	0.01	−0.01	0.03
Climate change attitudes (Kotcher)	696	−0.03	0.02	−0.07	0.01
Belief superiority (Raimi and Hall)	393	−0.01	0.04	−0.09	0.07
Belief extremity (Raimi and Hall)	393	0.00	0.03	−0.07	0.06
Affective polarization (Skytte)	204	−0.02	0.05	−0.12	0.07
Outparty intolerance (Skytte)	204	−0.08	0.08	−0.23	0.08
Outparty intolerance 2 (Skytte)	203	0.00	0.08	−0.16	0.16
News evaluations (Climate change hoax condition, Zhao)	235	0.02	0.06	−0.10	0.14
News evaluations (Climate change real condition, Zhao)	220	0.02	0.06	−0.10	0.14

Table 2b
Coefficients for Self-Affirmation X Subgroup Indicator Interaction (published)

Outcome measure	Sample size	B	SE	CI lower	CI upper
Affective polarization (Levendusky)	1334	−0.01	0.01	−0.03	0.01
Discussion polarization (Levendusky)	1338	−0.02	0.01	−0.05	0.00
Outparty danger (Levendusky)	1345	−0.03	0.02	−0.06	0.00
Social distance (Levendusky)	1333	0.02	0.02	−0.01	0.05
Outparty blame (Lyons)	130	−0.04	0.09	−0.22	0.14
Risk perception (Lyons)	274	0.00	0.04	−0.08	0.08
ISAF insurgent attacks (Graph condition, Nyhan and Reifler)	525	0.00	0.02	−0.04	0.05
Jobs (Obama approval moderator, graph condition, Nyhan and Reifler)	247	0.02	0.06	−0.10	0.13
Jobs (Economy MIP moderator, graph condition, Nyhan and Reifler)	247	0.01	0.06	−0.10	0.12
Anthropogenic GW (Text condition, Nyhan and Reifler)	122	−0.05	0.17	−0.38	0.28
Anthropogenic GW (Graph condition, Nyhan and Reifler)	122	−0.31	0.15	−0.62	−0.01
Temperature trend (Text condition, Nyhan and Reifler)	122	−0.03	0.08	−0.19	0.13
Temperature trend (Graph condition, Nyhan and Reifler)	121	−0.06	0.08	−0.23	0.10

Discussion

Can self-affirmation mitigate the psychological effects of partisanship and polarization? This article pools a number of studies undertaken by independent research teams to test this claim. Across studies, contexts, and outcomes of interests, we found little evidence that self-affirmation manipulations have ameliorative effects.

Prior work in the self-affirmation literature reveals that such effects may be contingent on a number of factors (e.g. Borman et al. 2018; Ferrer and Cohen, 2019). For this reason, our analyses focus on moderation effects, as self-affirmation effects should be found for those who care more about politics and/or those most threatened by inconvenient claims or by interactions with their political opponents. In this collection of studies, we offer evidence of a number of scenarios in which self-affirmation does not seem to improve outcomes, even among these subgroups.

These findings should be understood in the context of their limitations. The previously unpublished studies we report draw on samples of varying size – though some are large ($N = 3{,}799$) others instead rely on sample sizes resembling those in the existing self-affirmation literature, which are underpowered to detect interaction effects (Blake and Gangestad, 2020). In other words, these findings should also be considered in the context of prior work that motivated these studies. While all teams were excited by the potential of self-affirmation to address phenomena with strong normative implications – misperceptions, conspiracy belief, and polarization among them – the relatively limited evidentiary value provided by the small sample sizes of and inconsistent methods of prior studies (Table 3) is clearer in hindsight. These prior studies often show contingent effects, though little work has directly replicated the specific contingencies, which vary from study to study. At the same time, the studies presented here may be missing key features necessary for self-affirmation to work. If the work presented here can help clarify the limits of self-affirmation, we believe that would be a valuable contribution.[4]

Kotcher (2016, p. 68–69) provides a potential empirical explanation for the self-affirmation method's inconsistent effects:

> "self-affirmation not only can activate multiple psychological processes, but [. . .] it may activate both productive and counter-productive processes simultaneously . . . This suggests that self-affirmation is at the same time both more complex than previously understood [. . .] and less precise as a potential intervention than one might hope."

Given this richness and complexity, self-affirmation may remain a topic worthy of study for its own sake. At the same time, self-affirmation interventions may not be precise enough to allow researchers to produce normatively desirable outcomes without the danger of concurrently eliciting negative responses. Or, the true effects of self-affirming people may be null or artifactual. Indeed, there is evidence of publication bias in other literatures applying the self-affirmation framework (Weisz et al. 2016; see also Protzko and Aronson, 2016). We hope the findings presented

---

[4]Notably, recent works suggest that "heterogeneity and moderation of effects is unlikely for a zero average true effect size, but increasingly likely for larger average true effect size," (Olsson-Collentine et al., 2020). As such, the search for potential hidden moderators may be costlier than its contribution to the scientific record merits.

Table 3

Summary of Prior Studies Testing Self-Affirmation in Political Contexts

Authors/Study	Self-affirmation approach	Subject matter	Dependent variable(s)	Sample	Year	Place	Key estimate
Cohen et al. 2007 Study 1	Information following full cross of SA/self-threat X convictions/rationality prime	9/11	Openness to information	Students ($N = 43$)	–	USA	SA X salience: $F(1, 38) = 4.62$, $p = .038$.
Cohen et al. 2007 Study 2	Information between full cross of SA/self-threat X convictions prime	Abortion	Number of concessions made	Students ($N = 35$), pro-choice	–	USA	SA X salience: $F(1, 29) = 8.03$, $p = .008$
Cohen et al. 2007 Study 3	Information between full cross of SA/self-threat X commitment/compromise prime	Abortion	Number of acceptable restrictions on abortion access	Students ($N = 39$), pro-choice	–	USA	SA X salience: $F(1, 33) = 8.98$, $p = .005$
Binning et al. 2010	Information following randomized SA	US politics/ debate performance	Candidate evaluations	Convenience, other ($N = 110$), partisans	2008	USA	SA X party: $F(1, 105) = 5.77$, $p = .018$
Van Prooijen et al. 2013*	SA without information treatment	Climate change	Perceptions of pro-environmental behaviors	Students ($N = 90$)	–	UK	SA X worldview: $\beta = .32$, $p = .01$, $\beta = -0.49$, $p < 0.001$, $\beta = .23$, $p = .07$, $\beta = .15$, $p = .25$
Van Prooijen and Sparks, 2014	Information following randomized SA	Climate change	(a) acceptance of climate change risk (b) efficacy	Students ($N = 88$)	–	UK	SA X initial beliefs: $B = -0.52$, $p = .009$, $B = -0.61$, $p = .007$
Binning et al. 2015 Study 1	Information following full cross of SA and information treatment	Obama/US politics	Approval ratings	Convenience, other ($N = 115$), Democrats	2009	USA	SA X normative info: $F(1, 111) = 3.93$, $p = .050$.

(*Continued*)

Table 3
(*Continued*)

Authors/Study	Self-affirmation approach	Subject matter	Dependent variable(s)	Sample	Year	Place	Key estimate
Binning et al. 2015 Study 2	Full cross of SA and four-cell information treatment (2 X 2 X 2)	Obama/US politics	Approval ratings	Convenience, other ($N = 159$), Republicans	2010	USA	SA X direction X data type: $F(1, 151) = 3.99$, $p = .047$
Binning et al. 2015 Study 3	Full cross of SA and information treatment	Obama/US politics	Approval ratings	MTurk ($N = 224$)	2012	USA	SA X article X American ID: $B = -.76$, $p = .043$
Badea et al. 2017	SA (six-cell: self/group/control X left/right values) without information treatment	Refugees	Behavioral intent toward refugees	Students ($N = 192$)	–	France	SA X value-type X congruence: $B = -0.44$, $p = .03$
Badea et al. 2018	SA (three-cell: self/group/control) without information treatment	Terrorism	Perceived threat (symbolic, realistic), support for discriminatory policy, prejudice	Students ($N = 240$)	–	France	SA X individualism: $\beta = -0.16$, $p < 0.05$, $\beta = -0.20$, $p < 0.01$, $\beta = -0.09$, n.s., $\beta = -0.24$, $p < 0.001$.
Carnahan et al. 2018 Study 1	Information following randomized SA	Common core (education policy)	Change in belief accuracy	SSI ($N = 301$)	2016	USA	SA X attitude: $B = -0.12$, $p < 0.05$
Carnahan et al. 2018 Study 2	Full cross of SA and information treatment	GMOs	Change in belief accuracy	MTurk ($N = 509$)	2017	USA	SA X attitude X correction: $B = -.16$, $p < 0.05$

Note: Van Prooijen et al. 2013 find significant, *polarizing* effects of SA, causing "more constructive pro-environmental motives among participants with positive ecological worldviews but led to less constructive pro-environmental motives among participants with negative ecological worldviews."

here help form a more complete picture of self-affirmation effects on an assortment of identity-driven cognitive and affective phenomena.

With these null results, we also offer a number of research design suggestions for those who wish to continue exploring whether self-affirmation interventions can produce normatively desirable outcomes related to politics. We encourage researchers to use preregistered designs that employ samples large enough to detect the interaction effects that this theory proposes (Blake and Gangestad, 2020) or to precisely estimate null results. The results we report rely on conceptual replications. Direct replications (with samples large enough to reliably detect interaction effects) may help identify whether the eight separate research teams whose work is presented here simply erred in how they applied self-affirmation, or whether the approach is less useful than previously thought.

Others might wish to advance work summarized in Tables 1a, 1b and 3 by manipulating identity salience or threat directly. The *experimental manipulation* of threat may be key to uncovering self-affirmation effects (Ferrer and Cohen, 2019). Indeed, we see this as an unclear proposition given the present state of the literature. While Cohen et al. (2007) manipulate salience, other prior studies (Binning et al., Badea et al., van Prooijen et al., and Carnahan et al.) do not. Moreover, our view is that none of these studies manipulate *threat* to identity. Consequently, it is hard to establish that any positive effects of a self-affirmation intervention are contingent on salience or threat manipulations. Cohen et al. (2007) complicate matters by comparing self-affirmation not to a control, but to a "self-threat" condition where participants report a time they failed to live up to an important value. Experimental paradigms that can reliably induce identity threat may help identify when and where self-affirmation interventions outside the lab may be effective.

Typically, self-affirmation research does not employ manipulation checks for the same reason studies manipulating self-esteem do not: the manipulation check itself may prime the intended state (McQueen and Klein, 2006). The development of manipulation checks that can evaluate whether a self-affirmation treatment is working as intended (without doubly treating respondents) would be a significant advance. Without manipulation checks, it is hard to know whether it is one's theory that has failed, or simply one's procedures that have failed (though see Fayant et al. 2017 and Hauser et al. 2018 on the limitations of manipulation checks more generally).

Conclusion

In this paper, we present a number of studies that attempted to apply the self-affirmation paradigm to political topics. We bring together the work of 8 separate research teams who conducted 11 separate experiments. These different research teams were all motivated by the promising results of studies presented in Table 3 to use self-affirmation as a tool to, for lack a better word, improve politics in some way. Unfortunately, self-affirmation did not generate results consistent with the theory in any of the 11 different experiments.

Supplementary Material. To view supplementary material for this article, please visit https://doi.org/10.1017/XPS.2020.46.

References

Ansolabehere, S., J. Rodden and J. M. Snyder. 2008. The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting. *American Political Science Review* 102(2): 215–32.

Badea, C., K. Binning, J. F. Verlhiac and D. K. Sherman. 2018. In the Aftermath of Terrorism: Effects of Self Versus Group Affirmation on Support for Discriminatory Policies. *Journal of Experimental Social Psychology* 76: 421–28.

Badea, C., J. L. Tavani, M. Rubin and T. Meyer. 2017. Self-affirmation, Political Value Congruence, and Support for Refugees. *Journal of Applied Social Psychology* 47(7): 355–65.

Binning, K. R., C. Brick, G. L. Cohen and D. K. Sherman. 2015. Going along Versus Getting it Right: The Role of Self-Integrity in Political Conformity. *Journal of Experimental Social Psychology* 56: 73–88.

Binning, K. R., D. K. Sherman, G. L. Cohen and K. Heitland. 2010. Seeing the Other Side: Reducing Political Partisanship via Self-affirmation in the 2008 Presidential Election. *Analyses of Social Issues and Public Policy* 10(1): 276–92.

Blake, K. R. and S. Gangestad. 2020. On Attenuated Interactions, Measurement Error, and Statistical Power: Guidelines for Social and Personality Psychologists. *Personality and Social Psychology Bulletin* 0146167220913363.

Boninger, D. S., J. A. Krosnick and M. K. Berent. 1995. Origins of Attitude Importance: Self-Interest, Social Identification, and Value Relevance. *Journal of Personality and Social Psychology* 68(1): 61.

Borman, G. D., J. Grigg, C. S. Rozek, P. Hanselman and N. A. Dewey. 2018. Self-Affirmation Effects are Produced by School Context, Student Engagement with the Intervention, and Time: Lessons from a District-Wide Implementation. *Psychological Science* 29(11): 1773–84.

Carnahan, D., Q. Hao, X. Jiang and H. Lee. 2018. Feeling Fine about Being Wrong: The Influence of Self-Affirmation on the Effectiveness of Corrective Information. *Human Communication Research* 44(3): 274–298.

Carpenter, C. J. 2020. Meta-Analyzing Apples and Oranges: How to Make Applesauce Instead of Fruit Salad. *Human Communication Research* 46(2–3): 322–33.

Cohen, G. L., J. Aronson and C. M. Steele. 2000. When Beliefs Yield to Evidence: Reducing Biased Evaluation by Affirming the Self. *Personality and Social Psychology Bulletin* 26(9): 1151–1164.

Cohen, G. L. and D. K. Sherman. 2014. The Psychology of Change: Self-Affirmation and Social Psychological Intervention. *Annual Review of Psychology* 65: 333–371.

Cohen, G. L., D. K. Sherman, A. Bastardi, L. Hsu, M. McGoey and L. Ross. 2007. Bridging the Partisan Divide: Self-Affirmation Reduces Ideological Closed-Mindedness and Inflexibility in Negotiation. *Journal of Personality and Social Psychology* 93(3): 415.

Correll, J., S. J. Spencer and M. P. Zanna. 2004. An Affirmed Self and an Open Mind: Self-Affirmation and Sensitivity to Argument Strength. *Journal of Experimental Social Psychology* 40(3): 350–356.

Fayant, M. P., H. Sigall, A. Lemonnier, E. Retsin and T. Alexopoulos. 2017. On the limitations of manipulation checks: An obstacle toward cumulative science. *International Review of Social Psychology* 30(1): 125–130. doi: http://doi.org/10.5334/irsp.102.

Ferrer, R. A. and G. L. Cohen. 2019. Reconceptualizing Self-Affirmation with the Trigger and Channel Framework: Lessons from the Health Domain. *Personality and Social Psychology Review* 23(3): 285–304.

Hanselman, P., C. S. Rozek, J. Grigg and G. D. Borman. 2017. New Evidence on Self-Affirmation Effects and Theorized Sources of Heterogeneity from Large-Scale Replications. *Journal of Educational Psychology* 109(3): 405.

Harris, P. R. and L. Napper. 2005. Self-Affirmation and the Biased Processing of Threatening Health-Risk Information. *Personality and Social Psychology Bulletin* 31(9): 1250–63.

Hauser, D. J., P. C. Ellsworth and R. Gonzalez. 2018. Are Manipulation Checks Necessary? *Frontiers in Psychology* 9: 998.

Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra and S. J. Westwood. 2019. The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science* 22: 129–46.

Kahan, D. M. and D. Braman. 2006. Cultural Cognition and Public Policy. *Yale L. & Pol'y Rev* 24: 149.

Klar, S. 2013. The Influence of Competing Identity Primes on Political Preferences. *The Journal of Politics* 75(4): 1108–24.

Kotcher, J. 2016. Reducing Resistance to Communication about Climate Change using an Applied Self-affirmation Intervention. Doctoral Dissertation, George Mason University. Retrieved from http://ebot.gmu.edu/bitstream/handle/1920/10545/Kotcher_gmu_0883E_11279.pdf?sequence=1&isAllowed=y

Lehmiller, J. J., A. T. Law and T. T. Tormala. 2010. The Effect of Self-Affirmation on Sexual Prejudice. *Journal of Experimental Social Psychology* 46(2): 276–85.

Levendusky, M. S. 2018. When Efforts to Depolarize the Electorate Fail. *Public Opinion Quarterly* 82(3): 583–92.

Lyons, B. 2018. Reducing Group Alignment in Factual Disputes? The Limited Effects of Social Identity Interventions. *Science Communication* 40(6): 789–807.

Lyons, B., C. Farhart, M. Hall, J. Kotcher, M. Levendusky, J. Miller, B. Nyhan, K. Raimi, J. Reifler, K. Saunders, R. Skytte and X. Zhao. 2020. Replication Data for: Self-Affirmation and Identity-Driven Political Behavior. *Harvard Dataverse*, V3. doi: 10.7910/DVN/HUJZMO.

Ma, Y., G. Dixon and J. D. Hmielowski. 2019. Psychological Reactance from Reading Basic Facts on Climate Change: The Role of Prior Views and Political Identification. *Environmental Communication* 13(1): 71–86.

Mason, L. 2018. *Uncivil Agreement: How Politics Became our Identity*. Chicago: University of Chicago Press.

McQueen, A. and W. M. Klein. 2006. Experimental manipulations of self-affirmation: A systematic review. *Self and Identity* 5(4): 289–354.

Miller, J. M., K. L. Saunders and C. E. Farhart. 2016. Conspiracy Endorsement as Motivated Reasoning: The Moderating Roles of Political Knowledge and Trust. *American Journal of Political Science* 60(4): 824–44.

Neyeloff, J. L., S. C. Fuchs and L. B. Moreira. 2012. Meta-Analyses and Forest Plots using a Microsoft Excel Spreadsheet: Step-by-Step Guide Focusing on Descriptive Data Analysis. *BMC Research Notes* 5(1): 52.

Nyhan, B. and J. Reifler. 2010. When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior* 32(2): 303–30.

Nyhan, B. and J. Reifler. 2019. The Roles of Information Deficits and Identity Threat in the Prevalence of Misperceptions. *Journal of Elections, Public Opinion and Parties* 29(2): 222–44.

Oliver, J. E. and T. J. Wood. 2014. Conspiracy Theories and the Paranoid Style (s) of Mass Opinion. *American Journal of Political Science* 58(4): 952–66.

Olsson-Collentine, A., J. M. Wicherts and M. A. L. M. van Assen. 2020. Heterogeneity in Direct Replications in Psychology and its Association with Effect Size. *Psychological Bulletin* 146(10): 922–40. doi: 10.1037/bul0000294.

Pingree, R. J., D. Brossard and D. M. McLeod. 2014. Effects of Journalistic Adjudication on Factual Beliefs, News Evaluations, Information Seeking, and Epistemic Political Efficacy. *Mass Communication and Society* 17(5): 615–38.

Protzko, J. and J. Aronson. 2016. Context Moderates Affirmation Effects on the Ethnic Achievement Gap. *Social Psychological and Personality Science* 7(6): 500–7.

Reavis, R. D., J. B. Ebbs, A. K. Onunkwo and L. M. Sage. 2017. A Self-Affirmation Exercise Does not Improve Intentions to Vaccinate among Parents with Negative Vaccine Attitudes (and may Decrease Intentions to Vaccinate). *PLoS One* 12(7): e0181368.

Saucier, D. A. and R. J. Webster. 2010. Social Vigilantism: Measuring Individual Differences in Belief Superiority and Resistance to Persuasion. *Personality and Social Psychology Bulletin* 36(1): 19–32.

Sherman, D. A., L. D. Nelson and C. M. Steele. 2000. Do Messages about Health Risks Threaten the Self? Increasing the Acceptance of Threatening Health Messages via Self-affirmation. *Personality and Social Psychology Bulletin* 26(9): 1046–58.

Sherman, D. K. and G. L. Cohen. 2006. The Psychology of Self-Defense: Self-Affirmation Theory. *Advances in Experimental Social Psychology* 38: 183–242.

Steele, C. M. 1988. The Psychology of Self-Affirmation: Sustaining the Integrity of the Self. *Advances in Experimental Social Psychology* 21: 261–302. Academic Press.

Taber, C. S. and M. Lodge. 2006. Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science* 50(3): 755–69.

Van Koningsbruggen, G. M., E. Das and D. R. Roskos-Ewoldsen. 2009. How Self-Affirmation Reduces Defensive Processing of Threatening Health Information: Evidence at the Implicit Level. *Health Psychology* 28(5): 563.

van Prooijen, A. M. and P. Sparks. 2014. Attenuating Initial Beliefs: Increasing the Acceptance of Anthropogenic Climate Change Information by Reflecting on Values. *Risk Analysis* 34(5): 929–36.

Van Prooijen, A. M., P. Sparks and D. C. Jessop. 2013. Promoting or Jeopardizing Lighter Carbon Footprints? Self-Affirmation can Polarize Environmental Orientations. *Social Psychological and Personality Science 4*(2): 238–43.

Weisz, B.M., Lawner, E.K., Quinn, D.M., & Johnson, B.T. (2016). *How Well do Self-Affirmation and Role Model Interventions Combat Stereotype Threat in Lab and Field Settings?* Society for Personality and Social Psychology annual meeting, Jan 28–30, in San Diego, CA.

Zaller, J. R. 1992. *The Nature and Origins Of Mass Opinion.* Cambridge: Cambridge University Press.